

(51) Int.Cl.⁶

H 0 4 N 7/15

識別記号

F I

H 0 4 N 7/15

審査請求 有 請求項の数17 O L (全 7 頁)

(21) 出願番号 特願平10-161583

(22) 出願日 平成10年(1998) 6 月10日

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 加村 ゆかり

東京都港区芝五丁目7番1号 日本電気株式会社内

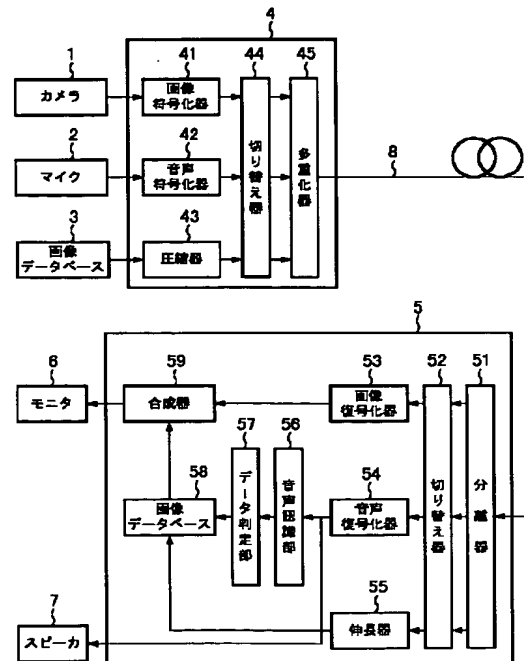
(74) 代理人 弁理士 京本 直樹 (外2名)

(54) 【発明の名称】 映像・音声通信装置と同装置を用いたテレビ会議装置

(57) 【要約】

【課題】 紙を用いずにタイムリーに補助画像情報を通常の映像に代えて、あるいは同時に表示できるようにする。

【解決手段】 映像と音声を送信する送信装置と、映像と音声を受信する受信部と、これらを再生する再生部とを有する受信装置とを備えた映像・音声通信装置であって、送信装置はさらに音声キーワードに対応する画像を送信する画像送信部を備え、受信装置は、画像を受信して蓄積する画像蓄積部と、受信した音声から音声キーワードを認識する音声認識部とを備えている。音声から音声キーワードが抽出されるとこの音声キーワードに対応する画像を画像蓄積部から選択して表示する。音声キーワードは、表示したい画像のみならず、表示の開始あるいは終了を指示するようにしてもよい。送信装置から表示制御信号を送信し、受信装置では指定された画像を画像蓄積部から選択して表示するようにしてもよい。



【特許請求の範囲】

【請求項 1】 映像と音声を送信する送信装置と、
前記映像と前記音声を受信する受信手段と、受信した映像と音声を再生する再生手段とを有する映像・音声受信装置とを備えた映像・音声通信装置であって、
前記映像・音声送信装置は、さらに、
画像を送信する画像送信手段と、
前記画像のうちから特定の画像を指定して該画像の表示を指示する表示制御信号を送信する表示制御信号送信手段とを備え、
前記映像・音声受信装置は、さらに、
前記画像を受信して蓄積する画像蓄積手段と、
前記表示制御信号を受信して指定された画像を前記画像蓄積手段から選択して前記映像に代えて表示する画像表示手段とを備えていることを特徴とする映像・音声通信装置。

【請求項 2】 前記画像は、前記映像の一部分に表示されることを特徴とする請求項 1 記載の映像・音声通信装置。

【請求項 3】 映像と音声を送信する送信装置と、
前記映像と前記音声を受信する受信手段と、受信した映像と音声を再生する再生手段とを有する映像・音声受信装置とを備えた映像・音声通信装置であって、
前記映像・音声送信装置は、さらに、
音声キーワードに対応する画像を送信する画像送信手段を備え、
前記映像・音声受信装置は、さらに、
前記画像を受信して蓄積する画像蓄積手段と、
受信した前記音声から前記キーワードを認識する音声認識手段とを備え、
前記音声キーワードが抽出されると、該音声キーワードに対応する前記画像を前記画像蓄積手段から選択して表示することを特徴とする映像・音声通信装置。

【請求項 4】 前記画像は、前記映像に代えて表示されることを特徴とする請求項 3 記載の映像・音声通信装置。

【請求項 5】 前記画像は、前記映像の一部分に表示されることを特徴とする請求項 3 記載の映像・音声通信装置。

【請求項 6】 前記音声キーワードは、
前記画像の表示の開始を指示する表示開始キーワードと、
前記画像の表示の終了を指示する表示終了キーワードとを含むことを特徴とする請求項 3 から請求項 5 までのいずれかの請求項に記載の映像・音声通信装置。

【請求項 7】 入力された映像を映像信号に変換して送信する映像信号送信手段と、
入力された音声を音声信号に変換して送信する音声信号送信手段とを有する映像・音声送信装置と、
前記映像信号を受信する映像信号受信手段と、

前記音声信号を受信する音声信号受信手段と、
受信された前記映像信号を映像に再生して映像表示部に表示する映像再生手段と、
受信された前記音声信号を音声に再生する音声再生手段と、を有する映像・音声受信装置とを備えた映像・音声伝送装置であって、
前記映像・音声送信装置は、さらに、
画像を前記映像・音声受信装置に向けて送信する画像送信手段と、

10 前記画像のうちから特定の画像を指定して該画像の表示を指示する表示制御信号を送信する表示制御信号送信手段とを備え、
前記映像・音声受信装置は、さらに、
前記画像と前記音声キーワードを受信して蓄積する画像データベースと、
15 前記表示制御信号を受信して指定された画像を前記画像蓄積手段から選択して前記映像表示部に表示される前記映像に代えて前記画像を表示することを特徴とする請求項 6 記載の映像・音声通信装置。

20 【請求項 8】 前記画像表示手段は、前記映像表示部の一部分に前記画像を表示することを特徴とする請求項 7 記載の映像・音声通信装置。

【請求項 9】 入力された映像を映像信号に変換して送信する映像信号送信手段と、
25 入力された音声を音声信号に変換して送信する音声信号送信手段とを有する映像・音声送信装置と、
前記映像信号を受信する映像信号受信手段と、
前記音声信号を受信する音声信号受信手段と、
受信された前記映像信号を映像に再生する映像再生手段と、
30 受信された前記音声信号を音声に再生する音声再生手段とを有する映像・音声受信装置とを備えた映像・音声伝送装置であって、
前記映像・音声送信装置は、さらに、

35 前記画像と前記音声キーワードを前記映像・音声受信装置に向けて送信する画像送信手段を備え、
前記映像・音声受信装置は、さらに、
前記画像と前記音声キーワードを受信して蓄積する画像データベースと、
40 前記音声信号から前記音声キーワードを認識する音声認識手段と、
前記音声キーワードが抽出された場合に該音声キーワードに対応する前記画像を前記画像データベースから選択して表示する画像表示手段とを備えていることを特徴とする映像・音声伝送装置。
45

【請求項 10】 前記映像再生手段は、前記映像を表示する映像表示部を備え、
前記画像表示手段は、前記映像表示部に表示される前記映像に代えて前記画像を表示することを特徴とする請求項 9 記載の映像・音声通信装置。
50

【請求項 1 1】 前記映像再生手段は、前記映像を表示する映像表示部を備え、前記画像表示手段は、前記映像表示部の一部分に前記画像を表示することを特徴とする請求項 9 記載の映像・音声通信装置。

【請求項 1 2】 前記音声認識手段は、さらに、前記画像の表示の開始を指示する開始指示キーワードと、前記画像の表示の終了を指示する終了指示キーワードとを備え、前記画像表示手段は、前記開始指示キーワードが抽出されたときに前記画像の表示を開始する画像表示開始手段と、前記終了指示キーワードが抽出されたときに前記画像の表示を終了する画像表示終了手段とを備えていることを特徴とする請求項 8 から請求項 1 1 までのいずれかの請求項に記載の映像・音声通信装置。

【請求項 1 3】 前記映像信号送信手段は、前記映像を符号化して映像符号化信号を出力する映像符号化器を含み、前記音声信号送信手段は、前記音声を符号化して音声符号化信号を出力する音声符号化器を含み、前記映像・音声送信装置は、さらに、前記画像を蓄積する送信側画像データベースと、前記画像を圧縮して圧縮画像信号を出力する圧縮器と、前記映像信号と前記音声信号と前記画像信号に代えて、前記映像符号化信号と前記音声符号化信号と前記圧縮画像信号とを多重化して多重化信号を送信する多重化器とを備え、前記映像・音声受信装置は、さらに、前記多重化信号を映像符号化信号と音声符号化信号と圧縮画像信号とにそれぞれ分離する分離器と、前記映像符号化信号を復号化して映像復号化信号を出力する映像復号化器と、前記音声符号化信号を復号化して音声復号化信号を出力する音声復号化器と、前記画像符号化信号を伸長して伸長画像信号を出力する伸長器とを備えていることを特徴とする請求項 7 または請求項 9 記載の映像・音声通信装置。

【請求項 1 4】 前記映像再生手段は、前記音声キーワードにより選定され画像データベースから取り出された画像信号と前記映像復号化信号とを合成する合成器を備えていることを特徴とする請求項 8 または請求項 1 1 記載の映像・音声通信装置。

【請求項 1 5】 前記映像信号送信手段は、前記映像を符号化して映像符号化信号を出力する映像符号化器を含み、前記音声信号送信手段は、前記音声を符号化して音声符号化信号を出力する音声符号化器を含み、前記映像・音声送信装置は、さらに、

前記画像を蓄積する送信側画像データベースと、前記画像を圧縮して圧縮画像信号を出力する圧縮器と、前記映像信号と前記音声信号と前記画像信号に代えて、前記映像符号化信号と前記音声符号化信号と前記圧縮画像信号とを多重化して多重化信号を送信する多重化器とを備え、前記映像・音声受信装置は、さらに、前記多重化信号を映像符号化信号と音声符号化信号と圧縮画像信号とにそれぞれ分離する分離器と、前記映像符号化信号を復号化して映像復号化信号を出力する映像復号化器と、前記音声符号化信号を復号化して音声復号化信号を出力する音声復号化器と、前記画像符号化信号を伸長して伸長画像信号を出力する伸長器とを備えていることを特徴とする請求項 1 0 から請求項 1 4 までのいずれかの請求項に記載の映像・音声通信装置。

【請求項 1 6】 前記画像は、静止画像または動画または文字データを含むことを特徴とする請求項 1 から請求項 1 5 までのいずれかの請求項に記載の映像・音声通信装置。

【請求項 1 7】 請求項 7 から請求項 1 6 までのいずれかの請求項に記載の映像・音声通信装置を備えたテレビ会議装置であって、前記映像信号送信手段は、映像を収録するカメラを含み、前記音声信号送信手段は、音声を収録するマイクを含み、前記映像再生手段は、前記映像を表示するモニタを含み、前記音声再生手段は、前記音声を出力するスピーカを含んでいることを特徴とするテレビ会議装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】本発明は、映像と音声を通信する映像・音声通信装置に関し、特に、映像・音声に加えてあらかじめ用意した画像も送信しこの画像を適宜相手方のモニタに表示できる通信装置であって、テレビ会議システムなどに適用されるものである。

【0 0 0 2】

【従来の技術】一般にテレビ会議を使用する場合、会議中相手の映像を見ていることが多いが、音声だけではわかりにくいような内容の場合、補足的に画像情報を用いることが会議の運用に有効な場合がある。例えば、会議の途中に図表などの資料を相手方に表示して説明したい場合などがある。

【0 0 0 3】

【発明が解決しようとする課題】しかし、このような場合、従来のテレビ会議においては、これらの補足情報を遠隔側から自由に選択することが困難なため、あらかじめ

めFAX等で補助情報を送付して、遠隔側からの指示で必要な補足情報を選択する方法がとられている。この方法では、どうしても紙の情報が補助的に必要となってしまう。また、相手方、すなわち補足情報に基づく説明を受ける側は、説明者側の指示をテレビ会議システムを通じて受けた上で該当する補足情報を選別して見る必要があり、手間がかかる上、映像画面からも目を離さなければならないなどの問題がある。

【0004】プレゼンテーション情報のリアルタイムな転送ができるようにするというだけののであれば、例えば、特開平6-169456号公報に記載されるようなプレゼンテーション会議システムが知られている。この会議システムは、編集部で作ったプレゼンテーション情報を会議の始まる前に、会議メンバの端末の記録部に転送して記録しておき、会議時には、所要の指示を送るだけで実行部にプレゼンテーションを行わせようというものである。

【0005】しかしながら、特開平6-169456号公報に記載のものは、あくまでも画像と音声情報からなるプレゼンテーション情報を予めプレゼンテーション端末側に送信しておき、プレゼンテーションを行うタイミングを遠隔から操作するにすぎず、テレビ会議などの途中で適宜に必要な画像情報などを映像に代えてモニタに表示するということができないという問題がある。

【0006】本発明の映像・音声通信装置は、紙を用いずに、タイムリーに補助画像情報を通常の映像に代えて、あるいは同時に表示できるようにすることを目的としている。

【0007】

【課題を解決するための手段】本発明の映像・音声通信装置は、上記問題を解決するために、映像と音声を送信する送信装置と、映像と音声を受信する受信部と、受信した映像と音声を再生する再生部とを有する受信装置とを備えた映像・音声通信装置であって、上記送信装置は、さらに画像を送信する画像送信部と、画像のうちから特定の画像を指定してこの画像の表示を指示する表示制御信号を送信する表示制御信号送信部とを備えている。一方、受信装置は、さらに画像を受信して蓄積する画像蓄積部と、送信装置からの表示制御信号を受信して指定された画像を画像蓄積部から選択して表示する画像表示部とを備えている。そして、映像・音声受信装置が表示制御信号を受信したときは、映像表示部に当該映像に代えて表示を指定された画像を表示するようにしたことを特徴としている。この表示は、映像に代えて映像表示部の全体に表示されるようにしても、また映像の一部に表示されるようにしてもよい。

【0008】本発明の映像・音声通信装置は、上述した画像を指定する表示制御信号を送信装置側で入力する代わりに、音声に基づいて表示が制御されるようにすることもできる。すなわち、送信装置は、上記基本構成に加

えて、音声キーワードに対応する画像を送信する画像送信部を備え、一方、受信装置は、画像を受信して蓄積する画像蓄積部と、受信した音声からキーワードを認識する音声認識部とを備えている。音声から音声キーワードが抽出されるとこの音声キーワードに対応する画像を画像蓄積部から選択して表示するようにしている。ここで、音声キーワードは、表示したい画像のみならず、表示の開始あるいは終了を指示するようにしてもよい。

【0009】本発明の映像・音声通信装置のより具体的な構成として、入力された映像を映像信号に変換して送信する映像信号送信部と、入力された音声を音声信号に変換して送信する音声信号送信部とを有する映像・音声送信装置と、映像信号を受信する映像信号受信部と、音声信号を受信する音声信号受信部と、受信された映像信号を映像に再生する映像再生部と、受信された音声信号を音声に再生する音声再生部とを有する映像・音声受信装置とを備えた映像・音声伝送装置を用いるようにする。そして、映像・音声送信装置は、さらに、画像を映像・音声受信装置に向けて送信する画像送信部と、画像のうちから特定の画像を指定してこの画像の表示を指示する表示制御信号を送信する表示制御信号送信部とを備えるようにしている。一方、映像・音声受信部は、さらに、画像と音声キーワードを受信して蓄積する画像データベースと、表示制御信号を受信して指定された画像を画像蓄積部から選択して表示する画像表示部とを備えている。

【0010】上述したように、受信装置側で表示されるべき画像は、送信側で表示制御信号を入力するのに代えて、音声信号から音声キーワードを認識する音声認識部を受信装置に備え、音声キーワードが抽出された場合にこの音声キーワードに対応する画像を画像データベースから選択して画像表示部に表示するようにしてもよい。

【0011】このような構成において、映像信号送信部は映像を符号化して映像符号化信号を出力する映像符号化器を有し、音声信号送信部は音声を符号化して音声符号化信号を出力する音声符号化器を有する。また、画像は圧縮器により圧縮された圧縮画像信号により送信されるようにする。さらに、映像、音声および画像は映像符号化信号と音声符号化信号と圧縮画像信号とを多重化して多重化信号により送信されるようにしてもよい。多重化信号は、受信装置側において、分離器により映像符号化信号、音声符号化信号、圧縮画像信号とにそれぞれ分離され、さらに、映像復号化器、音声復号化器および伸長器によりそれぞれもとの信号に再生される。

【0012】前記音声キーワードにより選定され画像データベースから取り出された画像信号と映像復号化信号とは合成器により合成され、映像表示部に同時に表示されるようにすることもできる。

【0013】以上のような映像・音声通信装置を用いて、映像信号送信部に映像を収録するカメラを、音声信

号送信部に音声を収録するマイクを、映像再生部に映像を表示するモニタを、音声再生部に音声を出力するスピーカを備えるようにし、テレビ会議装置を構成することもできる。

【0014】

【発明の実施の形態】次に、本発明の映像・音声通信装置と同装置を用いたテレビ会議装置について、図面を参照して以下に詳細に説明する。

【0015】図1は、本発明の映像・音声通信装置の一実施例の構成を示す図である。

【0016】図1に示されるように、本実施例では、理解を容易にするために、送信装置4を受信装置5と両者を接続する伝送路45から構成されているように図示されている。実際のテレビ会議装置や通信装置に適用される場合には、双方に送信装置と受信装置を備えるのが通常であるから、図1に示される本実施例の通信装置を双方向に配置すればよい。

【0017】図1において、送信装置4は、映像を取り込むカメラ1、音声を取り込むマイク2、および取り込まれた映像と音声をそれぞれ符号化する画像符号化器41、音声符号化器42を備えている。さらに、テレビ会議などで用いるための通常の映像とは別に相手方に見せたい画像を送信するための画像データベース3とここから取り込まれた情報を圧縮する圧縮器43を備えている。なお、画像データベース3には、スキャナなどによりその場で必要な画像を順次圧縮器43に入力されるようにしてもよいし、あらかじめすべての送信しておきたい画像情報を記録したものであってもよい。また、ここ

でいう画像は、静止画像のみならず動画像であってもよいし、また文章や表、グラフなどの文字情報を含むようなものであってもよい。

【0018】画像符号化器41、音声符号化器42、圧縮器43で得られた映像、音声および画像信号は切替器44を経て多重化器により多重化されて伝送路45を介して受信装置5に向けて送信される。本実施例では、補足的に用いようとしている画像情報についてはあらかじめ先に受信装置5に送信されるようにする。

【0019】受信装置5は、受信された多重化信号を映像信号、音声信号および画像信号にそれぞれ分離する分離器51と、切替器52と、画像復号化器53、音声復号化器54、伸長器55とを備えている。画像復号化器53、音声復号化器54、伸長器55でそれぞれもとの映像信号、音声信号、画像信号に再生される。

【0020】上述した先に送信された画像信号は、伸長器55で伸長された後、画像データベース58に蓄積される。なお、各画像信号は、後の送信装置側から送信される表示の指示に対応できるよう、識別子が付されている。すなわち、送信装置側では、あらかじめ会議等で使用する画像を識別子とともに画像データベース3に登録しておく。登録された画像信号は圧縮器43で圧縮され

て相手先に伝送され、伸長器55で伸長された後、相手先の画像データベース58へと登録される。

【0021】次に、会議等の実施中における映像の表示動作について説明する。会議中は、通常は送信装置4のカメラ1から撮影された映像信号が伝送され、受信装置5の側にあるモニタ6に表示される。また、送信装置4側の音声も同様に受信装置5の側に送信され、スピーカ7により再生される。

【0022】いま、送信装置4の側で、例えば説明のためにあらかじめ受信装置5側に登録して置いた画像を用いたい場合には、送信装置4側から表示したい画像を識別する上記識別子を含む表示制御信号を表示制御信号入力91から入力し送信する。この画像を特定するための識別子は、キーボードなどから入力されるようにすればよい。

【0023】表示制御信号を受信した受信装置5は、表示制御信号出力92に分離されここで抽出された表示制御信号により画像データベース58の中から対応する画像が選定されモニタ6に表示される。なお、この表示は、通常の映像に全く置き換えて表示してもよいし（映像とは別のモニタに表示する場合の当該モニタについては図示省略）、表示画面を分割して一部に表示をさせてもよい。一部に表示をする場合には、図1に示されるように、画像復号化器59からの映像信号と画像データベース58から取り出された画像信号とは合成器59により合成されることとなる。

【0024】次に、音声を用いたキーワードによる画像の識別を用いた表示について説明する。上記実施例においては、表示させたい画像は識別子を直接入力して特定することとしたが、本発明の映像・音声通信装置では、音声から画像を特定するためのキーワードを抽出し、このキーワードにより表示すべき画面を特定している点に特徴がある。

【0025】図2は、本発明の映像・音声通信装置の第2の実施例の構成を示すブロック図である。

【0026】送信装置4および受信装置5の構成は、図1に示される本発明の第1の実施例とほぼ同じである。但し、本実施例においては、音声復号化器54の後段に音声認識器56とデータ判別器57を備えている点が相違している。一方、送信装置4側には、表示させたい画像を特定するための識別子を入力するための機能は不要となる。

【0027】本実施例においては、補助画像はキーワードとなる音声とともにまず送信装置4の画像データベース3に登録される。登録されたデータは、圧縮器43で圧縮されて相手先に伝送され、伸長器55で伸長された後、相手先の画像データベース58へと登録される。また、キーワードは音声認識器56にも登録されて、画像データベースと関連づけられる。

【0028】カメラ1からの映像は、画像符号化器41

で符号化されて、切り替え器 4 4 を介して多重化器 4 5 へ供給される。マイク 2 からの音声は、音声符号化器 4 2 で符号化されて、切り替え器 4 4 を介して多重化器 4 5 へ供給される。切り替え器 4 4 では映像及び音声とデータとを切り替える。多重化器 4 5 では、前記符号化映像と前記符号化音声は、多重化器 4 5 により多重化されて伝送路 4 5 を経て受信装置 5 へと送信される。

【0029】受信装置 5 では、分離器 5 1 により伝送路 4 5 を介して送信された多重化信号は符号化映像と符号化音声とに分離され、さらに、符号化映像と符号化音声は切り替え器 5 2 へと送られる。符号化映像は、画像復号化器 5 3 へ供給され復号化され、合成器 5 9 へと供給される。一方、符号化音声は音声復号化器 5 4 へと供給され、復号化して音声認識器 5 6 とスピーカー 7 へと供給される。

【0030】ここで、音声認識器 5 6 では、入力された音声は認識され、あらかじめ登録されたキーワードと同じものが見つけられた場合には、その結果がデータ判定器 5 7 へ送られるように構成されている。データ判定器 5 7 では、認識された結果に対応するデータを画像データベース 5 8 から選択し、そのデータを合成器 5 9 に供給する。合成器 5 9 において、前記復号化画像と前記データとが合成され、モニタ 6 に送られる。モニタ 6 では、送信された映像と音声により選択されたデータとの合成画面が表示される。

【0031】上述した構成による通信装置を用いたキーワードによる画像の選択、表示についてもう少し具体的に説明する。

【0032】送信装置 4 側では、会議等のための発言された音声は符号化器 4 2 により符号化されて受信装置 5 側に送信される。いま、発言者が補助的に受信装置 5 のモニタに説明等のために補助的に画像を表示させたい場合には、画像を特定して表示させるためのあらかじめ定められた言葉を発する。例えば、複数の図面や写真があらかじめ受信装置 5 側の画像データベース 5 8 に登録されており、この中から図 5 をモニタ 6 に表示させたい場合には、「図 5 を表示」というキーワードを含む言葉、例えば「図 5 を表示しますのでご覧下さい。」という言葉はマイク 2 に向けて発する。

【0033】そうすると、音声信号は、受信装置 5 の音声復号化器 5 4 で復号化され、音声認識器 5 6 で「図 5」という画像を特定するキーワードが含まれているので、データ判別器 5 7 により画像データベース 5 8 にあらかじめ登録された画像の中から図 5 を選択する。そして、「表示」というキーワードも抽出されるので、当該画像がモニタ 6 に表示されることとなる。なお、この表示は、第 1 の実施例と同様、モニタ 6 に表示されている映像に代えて表示してもよいし、一部に表示してもよい。

【0034】上記キーワードが受信装置 5 側で抽出され

該当する画像の表示が開始された場合、例えば「終了」というような画像の表示を終了させるキーワードを登録しておき、この音声は抽出された場合に補助画像の表示が終了することとなり、もとの映像に自動的に戻される。

【0035】上述のように、画像を特定するキーワードの他に表示の開始と終了を識別して制御するキーワードを設けてもよいが、画像を特定するキーワードのみで表示が自動的に開始され一定時間表示の後、自動的にもとの映像画面に戻るようにしてもよいし、終了の場合のみ終了を指示するキーワードを持つようにしてもよい。

【0036】上記の本発明の第 2 の実施例の構成を採用することにより、画像データベースにキーワードと対になった補助映像データを蓄積しておき、受信した音声のキーワードに対応した補助映像データを選択し、受信した映像と合成して表示することによって、受け手側の操作なくタイムリーにデータを選択することができる。

【0037】以上説明した本発明の映像・音声通信装置は、送信装置と受信装置が一对になった通信装置を例に説明したが、通常はそれぞれの端末に送信装置と受信装置を一对にして双方向に映像・音声を通信できるようにしてテレビ会議が構成される。また、上記実施例では、発明の理解をより容易にするために、2 つの端末間のみの通信形態を例にあげたが、3 つ以上の端末間でテレビ会議装置を構成し、一つの送信装置から発せられた音声により画像表示のキーワードが 2 以上の受信装置で抽出され、各受信装置において、画像が表示されるようにすることができるのはいうまでもない。

【0038】

【発明の効果】以上説明したように、本発明の映像・音声通信装置は、あらかじめ画像を識別するための識別子を付して受信装置側に送信し、登録しておき、会議中に識別子を送信することにより受信装置側の画像データベースから当該画像を選択して表示するようにしているので、会議中は補助的に用いようとする画像信号の送信が不要になる。このため、受信側に瞬時に表示させたい画像を表示させることができるようになる。

【0039】さらに、本発明の映像・音声通信装置は、あらかじめキーワードと対になった補助映像を画像データベースに登録しておくことにより、受信した音声を音声認識により認識し、その音声に対応したデータを画像データベースより選択し、受信映像に代えて、あるいは合成して表示する構成を採用することもできる。この構成によれば、テレビ会議等において、送信側の音声に対応したデータを受信側の操作なく、音声と同時に受信側に表示することができるようになる。

【図面の簡単な説明】

【図 1】本発明の映像・音声通信装置の第 1 の実施例の構成を示すブロック図である。

【図 2】本発明の映像・音声通信装置の第 2 の実施例の

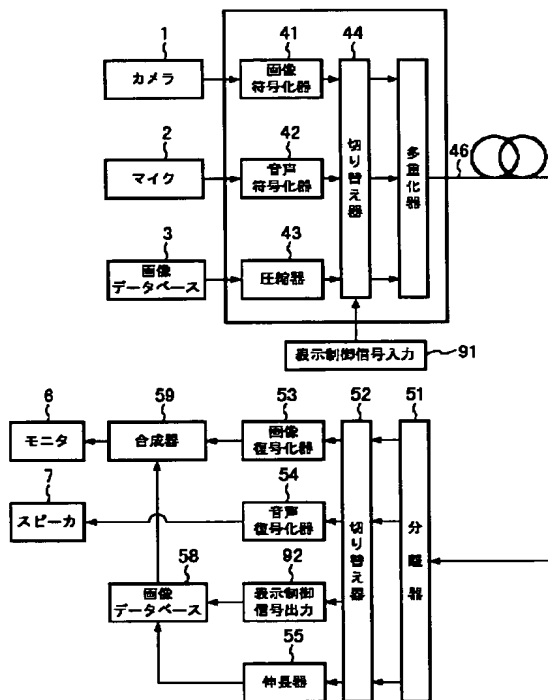
構成を示すブロック図である。

【符号の説明】

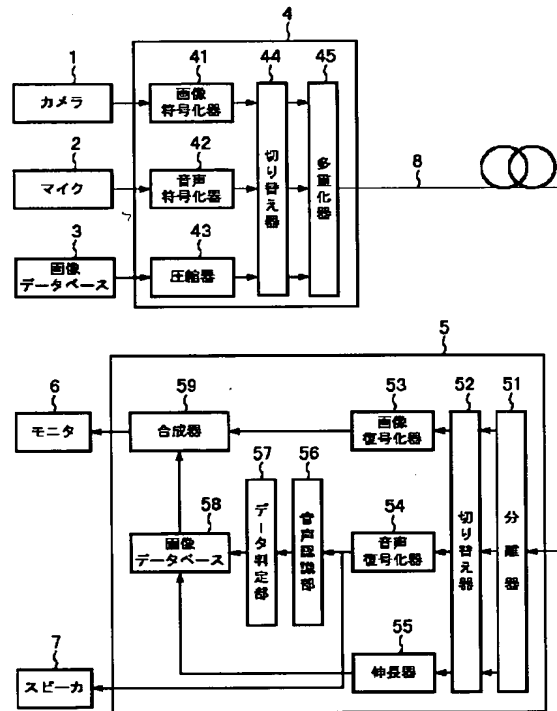
- 1 カメラ
- 2 マイク
- 3 画像データベース
- 4 CODECの送信部
- 4 1 画像符号化器
- 4 2 音声符号化器
- 4 3 圧縮器
- 4 4 切り替え器
- 4 5 多重化器
- 5 CODECの受信部
- 5 1 分離器

- 5 2 切り替え器
- 5 3 画像復号化器
- 5 4 音声復号化器
- 5 5 伸長器
- 05 5 6 音声認識器
- 5 7 データ判定器
- 5 8 画像データベース
- 5 9 合成器
- 6 モニタ
- 10 7 スピーカー
- 8 伝送路
- 9 1 表示制御信号入力
- 9 2 表示制御信号出力

【図 1】



【図 2】



PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-355747

(43)Date of publication of application : 24.12.1999

(51)Int.Cl.

H04N 7/15

(21)Application number : 10-161583

(71)Applicant : NEC CORP

(22)Date of filing : 10.06.1998

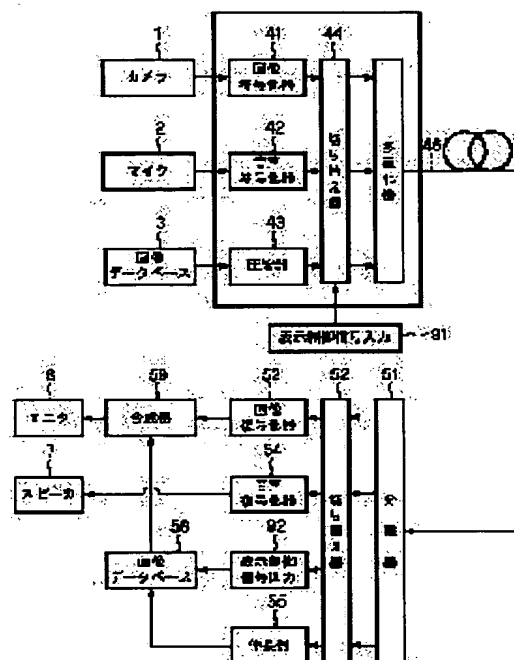
(72)Inventor : KAMURA YUKARI

(54) VIDEO/SOUND COMMUNICATION EQUIPMENT AND VIDEO CONFERENCE EQUIPMENT USING THE SAME EQUIPMENT

(57)Abstract:

PROBLEM TO BE SOLVED: To unnecessitate transmission of a picture signal to be subordinately used during a conference by preliminarily transmitting a picture to the receiver side by adding an identifier to it and registering it at the transmitter side, selecting a picture corresponding to the identifier from a picture database and displaying it at the receiver side when the identifier is transmitted during the conference.

SOLUTION: A video, a sound and a picture signal obtained by picture and sound encoders 41, 42 and a compressor 43 from a camera 1, a microphone 2 and a picture database 3 are processed, multiplexed and transmitted by the transmitter side. The received multiplexed signal is separated by a separator 51 and an original signal is outputted to a monitor 6, a speaker 7 and a picture database 58 via a switch 52 by picture and sound decoders 53, 54 and an extension device 55 at the receiver side. The picture signal is preliminarily transmitted to the reception side, and registered there by adding the identifier to it at display control signal input 91 by the transmission side and when the identifier is transmitted during the conference, the corresponding picture is displayed on the monitor 6. Thus, a subordinate picture is displayed without using paper.



LEGAL STATUS

[Date of request for examination] 10.06.1998

[Date of sending the examiner's decision of rejection] 02.07.2002

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The sending set which transmits an image and voice The image and voice receiving set which has the aforementioned image, a receiving means to receive the aforementioned voice, and the image which received and a reproduction means to reproduce voice They are an image and speech-communication equipment equipped with the above. the aforementioned image and voice sending set Furthermore, it has a display-control signal transmitting means to transmit the display-control signal which specifies [from] a specific picture to be a picture transmitting means to transmit a picture among the aforementioned pictures, and directs the display of this picture. The aforementioned image and voice receiving set are characterized by having further an image storage means to receive and accumulate the aforementioned picture, and an image display means to choose from the aforementioned image storage means the picture which received the aforementioned display-control signal and was specified, and to replace with and display on the aforementioned image.

[Claim 2] The aforementioned picture is the image and speech-communication equipment according to claim 1 characterized by what is displayed on a part of aforementioned image.

[Claim 3] The sending set which transmits an image and voice The image and voice receiving set which has the aforementioned image, a receiving means to receive the aforementioned voice, and the image which received and a reproduction means to reproduce voice They are an image and speech-communication equipment equipped with the above. the aforementioned image and voice sending set It has a picture transmitting means to transmit the picture corresponding to a voice keyword. furthermore, the aforementioned image and voice receiving set Furthermore, if it has an image storage means to receive and accumulate the aforementioned picture, and a speech recognition means to recognize the aforementioned keyword from the aforementioned voice which received and the aforementioned voice keyword is extracted, it will be characterized by choosing the aforementioned picture corresponding to this voice keyword from the aforementioned image storage means, and displaying it.

[Claim 4] The aforementioned picture is the image and speech-communication equipment according to claim 3 characterized by what is replaced with and displayed on the aforementioned image.

[Claim 5] The aforementioned picture is the image and speech-communication equipment according to claim 3 characterized by what is displayed on a part of aforementioned image.

[Claim 6] The aforementioned voice keyword is an image and speech-communication equipment given in one from the claim 3 characterized by including the display start keyword which directs the start of a display of the aforementioned picture, and the display end keyword which directs the end of a display of the aforementioned picture to a claim 5 of claims.

[Claim 7] A video-signal transmitting means to change the inputted image into a video signal and to transmit The image and voice sending set which has a sound signal transmitting means to change the inputted voice into a sound signal and to transmit A video-signal receiving means to receive the aforementioned video signal A sound signal receiving means to receive the aforementioned sound signal The image and voice receiving set which has an image-reproduction means to reproduce the received aforementioned video signal on an image, and to display on the

graphic display section, and a voice reproduction means to reproduce the received aforementioned sound signal to voice. They are an image and speech-communication equipment according to claim 6 equipped with the above. the aforementioned image and voice sending set. Furthermore, a picture transmitting means to turn a picture to the aforementioned image and voice receiving set, and to transmit. It has a display-control signal transmitting means to transmit the display-control signal which specifies [from] a specific picture among the aforementioned pictures, and directs the display of this picture. the aforementioned image and voice receiving set. Furthermore, it is characterized by replacing with the aforementioned image which chooses from the aforementioned image storage means the image database which receives and accumulates the aforementioned picture and the aforementioned voice keyword, and the picture which received the aforementioned display-control signal and was specified, and is displayed on the aforementioned graphic display section, and displaying the aforementioned picture.

[Claim 8] The aforementioned image display means is the image and speech-communication equipment according to claim 7 characterized by displaying the aforementioned picture on a part of aforementioned graphic display section.

[Claim 9] A video-signal transmitting means to change the inputted image into a video signal and to transmit. The image and voice sending set which has a sound signal transmitting means to change the inputted voice into a sound signal and to transmit. A video-signal receiving means to receive the aforementioned video signal. A sound signal receiving means to receive the aforementioned sound signal. The image and voice receiving set which has an image-reproduction means to reproduce the received aforementioned video signal on an image, and a voice reproduction means to reproduce the received aforementioned sound signal to voice. They are an image and voice transmission equipment equipped with the above. the aforementioned image and voice sending set. It has a picture transmitting means to turn the aforementioned picture and the aforementioned voice keyword to the aforementioned image and voice receiving set, and to transmit. furthermore, the aforementioned image and voice receiving set. Furthermore, the image database which receives and accumulates the aforementioned picture and the aforementioned voice keyword. It is characterized by having a speech recognition means to recognize the aforementioned voice keyword from the aforementioned sound signal, and an image display means to choose the aforementioned picture corresponding to this voice keyword from the aforementioned image database, and to display it when the aforementioned voice keyword is extracted.

[Claim 10] They are the image and speech-communication equipment according to claim 9 characterized by equipping the aforementioned image-reproduction means with the graphic display section which displays the aforementioned image, replacing the aforementioned image display means with the aforementioned image displayed on the aforementioned graphic display section, and displaying the aforementioned picture.

[Claim 11] They are the image and speech-communication equipment according to claim 9 which the aforementioned image-reproduction means is equipped with the graphic display section which displays the aforementioned image, and is characterized by the aforementioned image display means displaying the aforementioned picture on a part of aforementioned graphic display section.

[Claim 12] An image and speech-communication equipment given in one from a claim 8 to a claim 11 which is characterized by providing the following of claims. The aforementioned speech recognition means is a start directions keyword which directs the start of a display of the aforementioned picture further. It is an image display start means by which have the end directions keyword which directs the end of a display of the aforementioned picture, and the aforementioned image display means starts the display of the aforementioned picture when the aforementioned start directions keyword is extracted. An image display end means to end the display of the aforementioned picture when the aforementioned end directions keyword is extracted.

[Claim 13] The image and speech-communication equipment according to claim 7 or 9 characterized by providing the following. The aforementioned image and voice sending set are a

transmitting-side image database which accumulates the aforementioned picture further including the voice coding machine which the aforementioned sound signal transmitting means encodes the aforementioned voice, and outputs a voice coded signal including the image coding machine which the aforementioned video-signal transmitting means encodes the aforementioned image, and outputs an image coded signal. The compressor which compresses the aforementioned picture and outputs a compression picture signal It is the eliminator with which it replaces with the aforementioned video signal, the aforementioned sound signal, and the aforementioned picture signal, and has the multiplexing machine which multiplexes the aforementioned image coded signal, the aforementioned voice coded signal, and the aforementioned compression picture signal, and transmits a multiplexed signal, and the aforementioned image and voice receiving set divide the aforementioned multiplexed signal into an image coded signal, a voice coded signal, and a compression picture signal further, respectively. The image decryption machine which decrypts the aforementioned image coded signal and outputs an image decryption signal, the voice decryption machine which decrypts the aforementioned voice coded signal and outputs a voice decryption signal, and the expander which elongates the aforementioned picture coded signal and outputs an extension picture signal

[Claim 14] The aforementioned image-reproduction means is the image and speech-communication equipment according to claim 8 or 11 characterized by having the synthetic vessel which compounds the picture signal and the aforementioned image decryption signal which were selected by the aforementioned voice keyword and taken out from the image database.

[Claim 15] An image and speech-communication equipment given in one from a claim 10 to a claim 14 which is characterized by providing the following of claims The aforementioned image and voice sending set are a transmitting-side image database which accumulates the aforementioned picture further including the voice coding machine which the aforementioned sound signal transmitting means encodes the aforementioned voice, and outputs a voice coded signal including the image coding machine which the aforementioned video-signal transmitting means encodes the aforementioned image, and outputs an image coded signal. The compressor which compresses the aforementioned picture and outputs a compression picture signal It is the eliminator with which it replaces with the aforementioned video signal, the aforementioned sound signal, and the aforementioned picture signal, and has the multiplexing machine which multiplexes the aforementioned image coded signal, the aforementioned voice coded signal, and the aforementioned compression picture signal, and transmits a multiplexed signal, and the aforementioned image and voice receiving set divide the aforementioned multiplexed signal into an image coded signal, a voice coded signal, and a compression picture signal further, respectively. The image decryption machine which decrypts the aforementioned image coded signal and outputs an image decryption signal, the voice decryption machine which decrypts the aforementioned voice coded signal and outputs a voice decryption signal, and the expander which elongates the aforementioned picture coded signal and outputs an extension picture signal

[Claim 16] The aforementioned picture is an image and speech-communication equipment given in one from the claim 1 characterized by including a static image, a dynamic image, or an alphabetic data to a claim 15 of claims.

[Claim 17] The aforementioned voice reproduction means is TV-conference equipment characterized by to be included the loudspeaker which outputs the aforementioned voice including the monitor with which the aforementioned image-reproduction means displays the aforementioned image including the microphone with which voice is mentioned in the aforementioned sound signal transmitting means including the camera with which it is TV-conference equipment which equipped one from a claim 7 to a claim 16 of claims with the image and the speech-communication equipment of a publication, and an image is mentioned in the aforementioned video-signal transmitting means.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[The technical field to which invention belongs] this invention is the communication device which also transmits an image and the picture which prepared voice beforehand especially about the image and speech-communication equipment which communicates in addition to an image and voice, and can display this picture on the other party's monitor suitably, and is applied to a video conference system etc.

[0002]

[Description of the Prior Art] Although a partner's image is seen in many cases during the meeting when using a TV conference generally, in the case of a content which is unclear, it may be effective in employment of a meeting only with voice to use image information additionally. For example, there is a case where he wants to display and explain data, such as a chart, to the other party in the middle of a meeting etc.

[0003]

[Problem(s) to be Solved by the Invention] However, in such a case, at the conventional TV conference, auxiliary information is beforehand sent by FAX etc. and eye a difficult hatchet and the method of choosing required supplementary information with the directions from a remote side are taken for choosing these supplementary information from a remote side as freedom. By this method, the information on paper will surely be needed auxiliary. Moreover, the side given based on the other party, i.e., supplementary information, needs to sort out and look at the supplementary information which corresponds after receiving the directions by the side of an explainer through a video conference system, and has the problem of time and effort having to let an image screen out of sight this top.

[0004] If it can be made to perform a transfer [real time / information / presentation] /, a presentation conference system which is indicated by JP,6-169456,A is known, for example. This conference system transmits and records the presentation information made in the editorial department on the Records Department of the terminal of a meeting member, before a meeting starts, and it is a thing of making a presentation perform to a statement part only by sending necessary directions at the time of a meeting.

[0005] However, the thing given in JP,6-169456,A transmits beforehand the presentation information which consists of a picture and speech information to the last to the presentation terminal side, and does not pass over it to operate from remoteness the timing which performs a presentation, but has the problem that it cannot be said that a TV conference etc. replaces required image information etc. with an image suitably in the middle of, and displays on a monitor.

[0006] The image and speech-communication equipment of this invention aim at replacing auxiliary image information with the usual image timely, or enabling it to display simultaneously, without using paper.

[0007]

[Means for Solving the Problem] The sending set which transmits an image and voice in order that the image and speech-communication equipment of this invention may solve the above-

mentioned problem, They are an image and speech-communication equipment equipped with the receiving set which has an image, the receive section which receives voice, and the image which received and the reproduction section which reproduces voice. the above-mentioned sending set It has the display-control signal transmitting section which transmits the display-control signal which specifies [from] a specific picture among pictures to be the picture transmitting section which furthermore transmits a picture, and directs the display of this picture. On the other hand, the receiving set is equipped with the image storage section which receives and accumulates a picture further, and the image display section which chooses from the image storage section the picture which received the display-control signal from a sending set, and was specified, and displays it. And when an image and a voice receiving set receive a display-control signal, it is characterized by displaying the picture which it replaced [picture] with the image concerned and had the display specified by the graphic display section. Even if it replaces this display with an image and it is displayed on the whole graphic display section, it may be made to be displayed on a part of image.

[0008] Based on voice, a display can be controlled instead of the image and speech-communication equipment of this invention inputting the display-control signal which specifies the picture mentioned above by the sending set side. That is, in addition to the above-mentioned basic composition, the sending set was equipped with the picture transmitting section which transmits the picture corresponding to a voice keyword, and, on the other hand, the receiving set is equipped with the image storage section which receives and accumulates a picture, and the speech recognition section which recognizes a keyword from the voice which received. If a voice keyword is extracted from voice, the picture corresponding to this voice keyword is chosen from the image storage section, and it is made to display it. You may make it a voice keyword direct a start or end of not only a picture but a display to display here.

[0009] The video-signal transmitting section which changes the inputted image into a video signal as more concrete composition of the image and speech-communication equipment of this invention, and transmits, The image and voice sending set which has the sound signal transmitting section which changes the inputted voice into a sound signal and transmits, An image and voice transmission equipment equipped with the image and the voice receiving set which has the video-signal receive section which receives a video signal, the sound signal receive section which receives a sound signal, the image-reproduction section which reproduces the received video signal on an image, and the voice reproduction section which reproduces the received sound signal to voice are used. And it is made for an image and a voice sending set to be equipped with the display-control signal transmitting section which transmits the display-control signal which specifies [from] a specific picture further to be the picture transmitting section which turns a picture to an image and a voice receiving set, and transmits among pictures, and directs the display of this picture. On the other hand, the image and the voice receive section are equipped with the image display section which chooses from the image storage section the image database which receives and accumulates a picture and a voice keyword, and the picture which received the display-control signal and was specified, and displays it further.

[0010] When it replaces with inputting a display-control signal by the transmitting side, a receiving set is equipped with the speech recognition section which recognizes a voice keyword from a sound signal and a voice keyword is extracted, the picture which should be displayed by the receiving set side chooses the picture corresponding to this voice keyword from an image database, and you may make it display it on the image display section, as mentioned above.

[0011] In such composition, the video-signal transmitting section has the image coding machine which encodes an image and outputs an image coded signal, and the sound signal transmitting section has the voice coding machine which encodes voice and outputs a voice coded signal. Moreover, a picture is transmitted according to the compression picture signal compressed by the compressor. Furthermore, an image, voice, and a picture multiplex an image coded signal, a voice coded signal, and a compression picture signal, and may be made to be transmitted by the multiplexed signal. It is separated into an image coded signal, a voice coded signal, and a compression picture signal by the eliminator at a receiving set side, respectively, and a

multiplexed signal is further reproduced by the signal of a basis by the image decryption machine, the voice decryption machine, and the expander, respectively.

[0012] The picture signal and image decryption signal which were selected by the aforementioned voice keyword and taken out from the image database are compounded with a synthetic vessel, and can be simultaneously displayed on the graphic display section.

[0013] It has the loudspeaker which outputs voice for the monitor which displays an image for the microphone which records voice for the camera which records an image on the video-signal transmitting section on the sound signal transmitting section on the image-reproduction section to the voice reproduction section using the above image and speech-communication equipments, and TV conference equipment can also be constituted.

[0014]

[Embodiments of the Invention] Next, the TV conference equipment using the image and the speech-communication equipment, and this equipment of this invention is explained in detail below with reference to a drawing.

[0015] Drawing 1 is drawing showing the composition of one example of the image and speech-communication equipment of this invention.

[0016] As shown in drawing 1, in order to make an understanding easy, by this example, it is illustrated as the sending set 4 is constituted from a transmission line 45 which connects both with a receiving set 5. What is necessary is just to arrange bidirectionally the communication device of this example shown in drawing 1, since both sides are usually equipped with a sending set and a receiving set when applied to actual TV conference equipment and an actual communication device.

[0017] The sending set 4 is equipped with the picture coding machine 41 and the voice coding machine 42 which encode the camera 1 which incorporates an image, the microphone 2 which incorporates voice and the incorporated image, and voice, respectively in drawing 1. Furthermore, it has the compressor 43 which compresses the information incorporated from the image database 3 for transmitting a picture showing the other party apart from the usual image for using at a TV conference etc., and here. In addition, a required picture may be made to be inputted on that spot with a scanner etc. by the compressor 43 one by one, and all image information to transmit may be beforehand recorded on an image database 3. A picture here may be not only a static image but a dynamic image, and seems moreover, to include alphabetic information, such as a text, and a table, a graph.

[0018] The image, the voice, and the picture signal which were acquired by the picture coding machine 41, the voice coding machine 42, and the compressor 43 are multiplexed with a multiplexing vessel through a switcher 44, and are transmitted towards a receiving set 5 through a transmission line 45. It is made to be beforehand transmitted to a receiving set 5 previously in this example about the image information which it is going to use additionally.

[0019] The receiving set 5 is equipped with the eliminator 51 which divides the received multiplexed signal into a video signal, a sound signal, and a picture signal, respectively, the switcher 52, and the picture decryption machine 53, the voice decryption machine 54 and an expander 55. The video signal of a basis, a sound signal, and a picture signal are reproduced, respectively by the picture decryption machine 53, the voice decryption machine 54, and the expander 55.

[0020] The picture signal which was mentioned above and which was transmitted previously is accumulated at an image database 58, after being elongated by the expander 55. In addition, the identifier is attached so that each picture signal can respond to directions of the display transmitted from a next sending set side. That is, in the sending set side, the picture beforehand used at a meeting etc. is registered into the image database 3 with the identifier. The registered picture signal is registered to the image database 58 of the partner point, after being compressed by the compressor 43, being transmitted to the partner point and elongated by the expander 55.

[0021] Next, the display action of the image under implementation of a meeting etc. is explained. During a meeting, the video signal usually photoed from the camera 1 of a sending set 4 is transmitted, and it is displayed on the monitor 6 in a receiving set 5 side. Moreover, it is similarly

transmitted to a receiving set 5 side, and the voice by the side of a sending set 4 is also reproduced by the loudspeaker 7.

[0022] The display-control signal containing the above-mentioned identifier which discriminates a picture to display from a sending set 4 side is inputted from the display-control signal input 91 and it transmits to use now the picture beforehand registered and put on the receiving set 5 side by the sending set 4 side for explanation. The identifier for specifying this picture should just be inputted from a keyboard etc.

[0023] The picture which corresponds out of an image database 58 with the display-control signal which was divided into the display-control signal output 92, and was extracted here is selected, and the receiving set 5 which received the display-control signal is displayed on a monitor 6. In addition, this display may completely be transposed to the usual image, and may be displayed on it, and the display screen may be divided (about the monitor concerned in the case of displaying on a monitor different from an image, it is an illustration abbreviation), and you may make it display in part. When displaying in part, as shown in drawing 1, the video signal from the picture decryption machine 59 and the picture signal taken out from the image database 58 will be compounded with the synthetic vessel 59.

[0024] Next, the display using discernment of the picture by the keyword using voice is explained. Although the picture which you want to display carries out the direct input of the identifier and being specified in the above-mentioned example, with the image and speech-communication equipment of this invention, the keyword for specifying a picture is extracted from voice, and the feature is in the point of specifying the screen which should be displayed by this keyword.

[0025] Drawing 2 is the block diagram showing the composition of the 2nd example of the image and speech-communication equipment of this invention.

[0026] The composition of a sending set 4 and a receiving set 5 is almost the same as the 1st example of this invention shown in drawing 1. However, in this example, the point which equips the latter part of the voice decryption machine 54 with the speech recognition machine 56 and the data distinction machine 57 is different. On the other hand, the function for inputting the identifier for specifying the picture which you want to display as a sending set 4 side becomes unnecessary.

[0027] In this example, an auxiliary picture is first registered into the image database 3 of a sending set 4 with the voice used as a keyword. The registered data are registered to the image database 58 of the partner point, after being compressed by the compressor 43, being transmitted to the partner point and elongated by the expander 55. Moreover, a keyword is registered also into the speech recognition machine 56, and is related with an image database.

[0028] It encodes with the picture coding vessel 41, and the image from a camera 1 is supplied to the multiplexing machine 45 through the change machine 44. It encodes with the voice coding vessel 42, and the voice from a microphone 2 is supplied to the multiplexing machine 45 through the change machine 44. An image and voice, and data are changed with the change vessel 44. With the multiplexing vessel 45, the aforementioned coding image and the aforementioned coding voice are multiplexed with the multiplexing vessel 45, and are transmitted to a receiving set 5 through a transmission line 45.

[0029] In a receiving set 5, the multiplexed signal transmitted through the transmission line 45 by the eliminator 51 is divided into a coding image and coding voice, and a coding image and coding voice are further sent to the change machine 52. A coding image is supplied to the picture decryption machine 53, is decrypted, and is supplied to the synthetic vessel 59. On the other hand, coding voice is supplied to the voice decryption machine 54, is decrypted and is supplied to the speech recognition machine 56 and a loudspeaker 7.

[0030] Here, the inputted voice is recognized, and when the same thing as the keyword registered beforehand is found, it consists of speech recognition machines 56 so that the result may be sent to the data judging machine 57. With the data judging vessel 57, the data corresponding to the recognized result are chosen from an image database 58, and the data is supplied to the synthetic vessel 59. In the synthetic vessel 59, the aforementioned decryption picture and the aforementioned data are compounded and it is sent to a monitor 6. In a monitor

6, the synthetic screen of the transmitted image and the data chosen by voice is displayed.

[0031] It explains a little more concretely about selection of the picture by the keyword using the communication device by the composition mentioned above, and a display.

[0032] In a sending set 4 side, it encodes with the coding vessel 42 and the voice said for the meeting etc. is transmitted to a receiving set 5 side. The language by means of which it was beforehand set for specifying and displaying a picture is emitted to display [speaker] a picture on the monitor of a receiving set 5 auxiliary because of explanation etc. auxiliary now. For example, two or more drawings and photographs are beforehand registered into the image database 58 by the side of a receiving set 5, and, for example, turn to a microphone 2 the word "have a look since drawing 5 is displayed", and it emits to display drawing 5 on a monitor 6 from this inside. [the language containing the keyword of "displaying drawing 5"]

[0033] If it does so, a sound signal is decrypted with the voice decryption vessel 54 of a receiving set 5, and since the keyword which specifies the picture of "drawing 5" with the speech recognition vessel 56 is contained, it will choose drawing 5 from the pictures beforehand registered into the image database 58 with the data distinction vessel 57. And since the keyword "a display" is also extracted, the picture concerned will be displayed on a monitor 6. In addition, this display may be replaced with and displayed on the image currently displayed on the monitor 6 like the 1st example, and may be displayed in part.

[0034] When the display of the picture to which the above-mentioned keyword is extracted by the receiving set 5 side, and corresponds is started (for example, when the keyword which terminates the display of a picture, such as "an end", is registered and this voice is extracted), the display of an auxiliary picture will be completed, and it is automatically returned to the image of a basis.

[0035] As mentioned above, although the keyword which discriminates and controls a start and end of a display to everything but the keyword which specifies a picture may be prepared, a display is automatically started only by the keyword which specifies a picture, and you may make it return to the image screen of a basis automatically, and may make it have the keyword which directs an end after a fixed time display only in an end.

[0036] By adopting the composition of the 2nd example of the above-mentioned this invention, data can be chosen timely without the operation by the side of a sink by accumulating the auxiliary image data which became a keyword and a pair to the image database, choosing the auxiliary image data corresponding to the keyword of the voice which received, compounding with the image which received, and displaying.

[0037] Although the image and speech-communication equipment of this invention explained above explained to the example the communication device from which the sending set and the receiving set became a couple, a TV conference is constituted, as a sending set and a receiving set are usually made each terminal at a couple and an image and voice can be communicated bidirectionally. Moreover, although the communication form only between two terminals was mentioned as the example in the above-mentioned example in order to make an understanding of invention easier, TV conference equipment is constituted among three or more terminals, the keyword of image display is extracted with two or more receiving sets by the voice uttered from one sending set, and it cannot be overemphasized in each receiving set that a picture can be displayed.

[0038]

[Effect of the Invention] Since the image and speech-communication equipment of this invention attach the identifier for discriminating a picture beforehand, and transmits and registers it into the receiving set side, and the picture concerned is chosen from the image database by the side of a receiving set by transmitting an identifier during a meeting, as explained above, and it is made to display, the transmission of a picture signal which it is going to use auxiliary becomes unnecessary during a meeting. For this reason, the picture which you want to display on a receiving side in an instant can be displayed now.

[0039] Furthermore, by registering into the image database the auxiliary image which became a keyword and a pair beforehand, the image and speech-communication equipment of this invention can recognize the voice which received by speech recognition, can choose the data

corresponding to the voice from an image database, and can replace them with a receiving image, or the composition compounded and displayed can also be used for it. According to this composition, at a TV conference etc., there is no operation of the data corresponding to the voice of a transmitting side of a receiving side, and it can display now on a receiving side simultaneously with voice.

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.*** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram showing the composition of the 1st example of the image and speech-communication equipment of this invention.

[Drawing 2] It is the block diagram showing the composition of the 2nd example of the image and speech-communication equipment of this invention.

[Description of Notations]

1 Camera

2 Microphone

3 Image Database

4 Transmitting Section of CODEC

41 Picture Coding Machine